학습과 예측의 유전 제어: 플라즈마 식각공정 데이터 모델링에의 응용

Genetic Control of Learning and Prediction: Application to Modeling of Plasma Etch Process Data

김 병 환*, 우 형 수, 곽 관 웅
(Byungwhan Kim, Hyung Soo Uh, and Kwan Woong Gwak)

Abstract : A technique to model plasma processes was presented. This was accomplished by combining the backpropagation neural network (BPNN) and genetic algorithm (GA). Particularly, the GA was used to optimize five training factor effects by balancing the training and test errors. The technique was evaluated with the plasma etch data, characterized by a face-centered Box Wilson experiment. The etch outputs modeled include Al etch rate, Al selectivity, DC bias, and silica profile angle. Scanning electron microscope was used to quantify the etch outputs. For comparison, the etch outputs were modeled in a conventional fashion. GA-BPNN models demonstrated a considerable improvement of more than 25% for all etch outputs only but he DC bias. About 40% improvements were even achieved for the profile angle and Al etch rate. The improvements demonstrate that the presented technique is effective to improving BPNN prediction performance.

Keywords : backpropagation neural network, genetic algorithm, plasma etching, model

I. Introduction

Plasmas play a crucial role in etching and depositing thin films. Prediction models of plasma processes are in demand for characterization, diagnosis, and control of plasma equipment. First principle models are subject to many simplifying assumptions due to the lack of understanding of physical and chemical processes. As an alternative, an adaptive neural network has been used to build a prediction model of plasma processes [1-3]. Compared to physical models, neural network models are advantageous in that they do not require any assumptions while producing quick predictions. Among the various neural network paradigms, the backpropagation neural network (BPNN) [4] has been the most widely adopted in plasma modeling. Constructing a BPNN model is complicated by the presence of several training factors, including the hidden neurons, training tolerance, initial weight distribution, and function gradients [5]. In most applications, training factor effects are typically optimized by experimentally tuning each factor individually. For a simultaneous optimization of training factors, a genetic algorithm (GA) [6] has been applied to model plasma process [3,5]. The GA-applied model just mentioned is somewhat limited in that only the training error was optimized. By balancing both training and test errors, a better prediction model is expected.

In this study, a BPNN model was constructed by controlling the training and test errors during a genetic learning. This technique was evaluated with the experimental data collected from the etching of silica thin films [7]. For a systematic modeling, the etch process was conduced according to a statistical experimental design. Optimized GA-BPNN models were compared to those constructed in conventional way.

* 책임저자(Corresponding Author)
논문접수 : 2007. 1. 25., 채택확정 : 2007. 2. 17.
김병환, 우형수 : 세종대학교 전자공학과
(kbwhan@sejong.ac.kr/hsuh@sejong.ac.kr)
곽관웅: 세종대학교 기계공학과
(kwgwak@sejong.ac.kr)
※ This work was supported by the Korean Research Foundation Grant funded by the Korean Government(MOEHRD) (KRF-2006-311-D00047).

II. Experimental Data

Schematic diagram of an inductively coupled plasma etch system is shown in Fig. 1. The process chamber serving as a ground plane is pumped by a turbomolecular pump, and the pressure is controlled by a down stream throttle valve. Gases are introduced via a multihole shower head and radio frequency (RF) bias power operating at 13.56 MHz is fed to the lower electrode with the diameter of 8.5 inches via a matching network. The upper chamber section was modified by a ceramic cylinder, through which RF source power is coupled via a multiturn helical coil operating at 2 MHz. The cylinder is closed at the upper end by a grounded electrode.

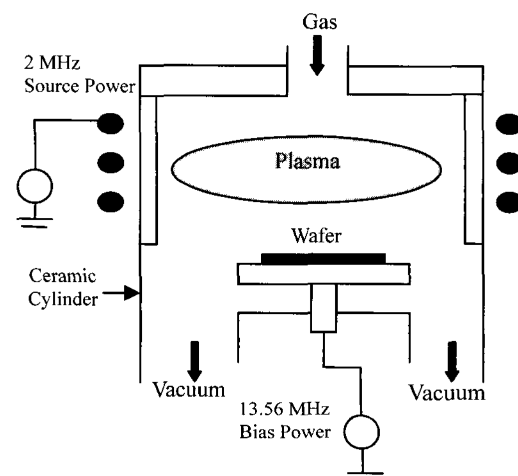


그림 1. 플라즈마 식각 시스템의 개요도.
Fig. 1. Schematic of plasma etch system.

Test patterns were fabricated on p-type silicon wafers of 5 inch diameter. A buffer-clad layer of about 25 um was deposited by the flame hydrolysis deposition method. The core layer of 8 um thickness was subsequently formed, which was formed as SiO_2-P_2O_5-B_2O_3-GeO_2. Here the percentages of P, B, and Ge were all less than 1%. After the evaporation of AlSi(1%) of about 400 nm, the waveguide was patterned using a contact aligner. The AlSi(1%) layer was first etched in a $BCl_3/Cl_2/CHF_3$ plasma using the Plasma Therm reactive ion etch system, where the RF power and pressure were set to 150 W and 10 mTorr, respectively. To remove the layer completely, the layer was subsequently 30 % overetched and the resist was then solvent-stripped. Following this, the silica core layer was etched in a CF_4/CHF_3 plasma using a Plasma Therm 690 ICP etch system.

The process parameters that were varied include a source power, bias power, and gas ratio. The total flow rate of gases, CHF_3 and CF_4, was set to 60 sccm and the flow rate of CHF_3 was varied from 10 to 50 sccm. The gas ratio was defined as the flow rate of CHF_3 divided by the flow rate of CF_4. To characterize the etch process, a 2^3 full factorial experiment [8] was used along with one center point. The experimental parameter ranges employed in the design are 100-800 W, 100-400 W, and 0.2-5.0, for the source power, bias power, and gas ratio, respectively. Additional six experiments were conducted to provide the test data for model evaluation. Consequently, a total of 15 experiments were conducted. The etch outputs modeled include a silica profile angle, aluminum (Al) etch rate, Al selectivity to silica, and DC bias. Using a Hitachi S800 scanning electron microscopy, the vertical etch rate of silica, R, was measured along with the vertical Al etch rate. The Al selectivity is defined as the silica etch rate divided by the Al etch rate. The profile angle (A) was estimated from the schematic of front view of etched silica film, depicted in Fig. 2. The A was defined as

$$A = \tan^{-1}[\frac{2R}{L-U}] \tag{1}$$

where U and L represent the widths of the original and etched patterns, respectively. The remaining DC bias was measured by a DC volt meter embedded in a RF match network.

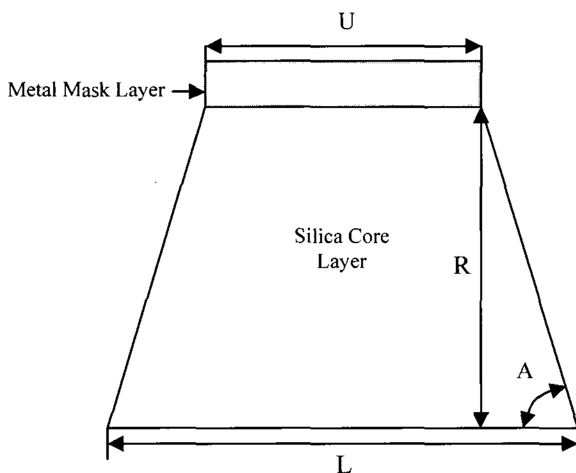


그림 2. 식각측정치의 개요도.

Fig. 2. Schematic for etch measurements.

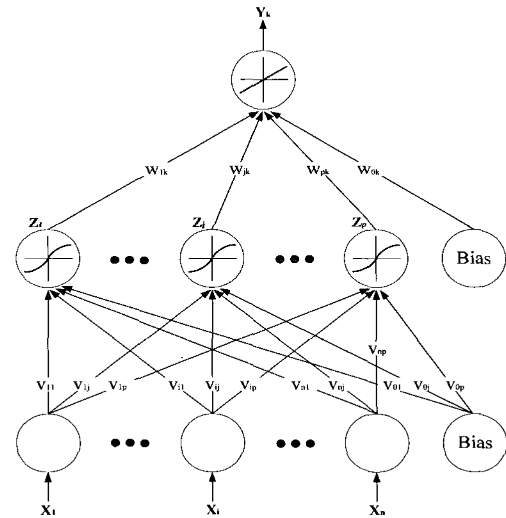


그림 3. BPNN의 개요도.

Fig. 3. Schematic of BPNN.

III. Neural Network

As shown in Fig. 3, the BPNN consists of three layers of neurons: input layer, hidden layer, and output layer. The number of hidden layers was set to unity. A hidden neuron in the hidden layer performs a nonlinear feature extraction on the data provided by the input and output layers. Depending on the number of hidden neurons (NHN), the BPNN prediction performance can vary significantly. The activation level (or firing strength) of a neuron in the hidden layer is determined by a bipolar sigmoid function denoted as

$$out_{i,k} = \frac{1-e(-\frac{in_{i,k}}{g_b})}{1+e(-\frac{in_{i,k}}{g_b})} \tag{2}$$

where $in_{i,k}$ and $out_{i,k}$ indicate the weighted input to the ith neuron in the kth layer and output from that neuron, respectively. The g_b represents the gradient of the bipolar sigmoid function. Meanwhile, the linear function adopted in the output layer is expressed as

$$out_{i,k} = in_{i,k} \cdot g_l \tag{3}$$

where g_l represents the gradient of the linear function. Apart from the three training factors (NHN, g_b, and g_l) already stated, the initial weight distribution (IWD) and the training tolerance (TT) influence the BPNN prediction considerably [5]. Hence, the total number of training factors to optimize is five. As a weight update scheme, meanwhile, the generalized delta rule [4] was adopted.

IV. Results

1. Conventional Model

BPNN models were constructed in conventional way. In other words, the effects of training factors on the BPNN prediction performance were optimized experimentally one by one. As an illustration, the Al etch rate was modeled. First, the effects of three training factors (TT, NHN, IWD) were optimized, followed by the optimization of two gradients. The experimental ranges for TT,

NHN, IWD employed in the first stage are 0.08-0.12, 2-4, and ±0.2-1.6, respectively. The prediction error was quantified with the root mean squared error (RMSE). For convenience, this error is called "P-RMSE". The results are shown in Table 1.

표 1. Al 식각률 모델의 예측에러.

Table 1. Prediction performance of Al etch model.

TT	P-RMSE (Å/min)	NHN	P-RMSE (Å/min)	IWD	P-RMSE (Å/min)
0.08	732	2	596	0.2	559
0.09	725	3	704	0.4	511
0.10	718	4	512	0.6	542
0.11	711			0.8	502
0.12	704			1.0	512
				1.2	513
				1.4	511
				1.6	508

표 2. Al 선택비 모델의 예측에러.

Table 2. Prediction performance of Al selectivity model.

TT	P-RMSE	NHN	P-RMSE	IWD	P-RMSE
0.08	2.76	2	2.45	0.2	2.30
0.09	2.73	3	2.67	0.4	2.32
0.10	2.71	4	3.69	0.6	2.42
0.11	2.68			0.8	2.42
0.12	2.66			1.0	2.45
				1.2	2.50
				1.4	2.56
				1.6	2.63

표 3. 프로파일 각 모델의 예측에러.

Table 3. Prediction performance of profile angle model.

TT	P-RMSE (Degree)	NHN	P-RMSE (Degree)	IWD	P-RMSE (Degree)
0.08	3.66	2	4.66	0.2	3.57
0.09	3.65	3	3.64	0.4	3.44
0.10	3.65	4	3.47	0.6	3.59
0.11	3.64			0.8	3.58
0.12	3.65			1.0	3.47
				1.2	3.53
				1.4	3.73
				1.6	4.02

표 4. DC bias 모델의 예측에러.

Table 4. Prediction performance of DC bias model.

TT	P-RMSE (V)	NHN	P-RMSE (V)	IWD	P-RMSE (V)
0.08	77.3	2	89.6	0.2	82.8
0.09	77.4	3	77.3	0.4	80.3
0.10	77.5	4	98.6	0.6	78.4
0.11	77.6			0.8	79.1
0.12	77.7			1.0	77.3
				1.2	75.2
				1.4	74.2
				1.6	76.8

From Table 1, one model with the smallest P-RMSE is obtained at 0.11 TT. It should be noted that in optimizing the TT effect the values of NHN and IWD were set to 3 and ±1, respectively. After setting the TT to 0.11, the NHN was then varied and one optimized model is obtained at 4 NHN. The corresponding P-RMSE is much smaller than that for the preceding model. This indicates that NHN affects the BPNN prediction accuracy considerably. Despite many variations in the IWD, the resulting P-RMSEs are no longer decreased appreciably. A slightly smaller RMSE is obtained only for the model constructed at ±1.6. In this way, the training factors were optimized and the results are shown in Tables 2-4 for the Al selectivity, profile angle, and DC bias, respectively.

In the second stage of optimization, the gradient effect was optimized while setting other training factors to their optimized values. For this, the g_b was increased from 0.4 to 2.0 with an increment of 0.2. For each g_b, the g_l was varied in the same way as g_b. Consequently, 81 combinations of function gradients were generated and the corresponding P-RMSEs are shown in Fig. 4. From Fig. 4, one best RMSE is obtained in the 10th case composed of 0.4 g_b and 0.8 g_l. The corresponding P-RMSE is 434 Å/min. Compared to the optimized P-RMSE determined in the first stage, this demonstrates an improvement of about 14.5%. This improvement reveals that the gradients affect the BPNN prediction performance considerably. Using the same process, other remaining three etch outputs were modeled and results are shown in Table 5.

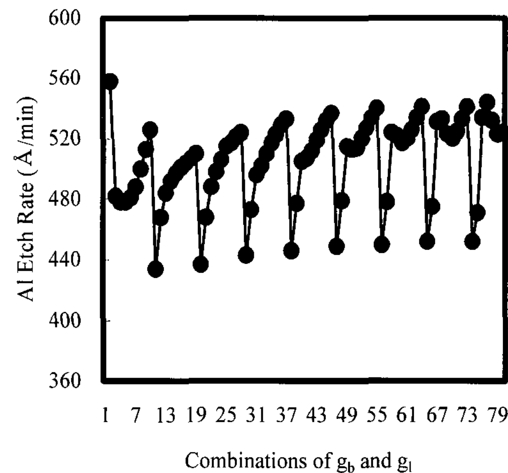


그림 4. 함수경사 조합의 함수로의 예측성능.

Fig. 4. Prediction performance as a function of combinations of function gradients.

표 5. 최적화된 학습인자와 식각모델의 예측성능.

Table 5. Optimized training factors and prediction performances of etch models.

Etch Outputs	TT	NHN	IWD	g_b	g_l	P-RMSE
Profile Angle (°)	0.11	4	0.4	0.8	0.4	2.85
Al Selectivity	0.12	2	0.2	1.2	1.2	2.26
DC Bias (V)	0.08	3	1.4	0.4	0.4	53.6
Al Etch Rate (Å/min)	0.12	4	0.8	0.4	0.8	434

표 6. GA 최적화에 이용된 학습인자의 실험적 범위.

Table 6. Experimental ranges of training factors employed in GA optimization.

Training Factors	Ranges
TT	0.08-0.12
NHN	3-6
IWD	0.2-1.6
g_b	0.4-2.0
g_b	0.4-2.0

표 7. Al 식각률에 대한 GA-BPNN 모델의 예측성능.

Table 7. Prediction performance of GA-BPNN model for the Al etch rate.

Q	P-RMSE (Å/min)	Fitness
0.1	261	0.0400
0.3	323	0.0037
0.5	341	0.0042
0.7	385	0.0058
0.9	458	0.0094

표 8. 최적화된 GA-BPNN 모델의 예측성능.

Table 8. Prediction performance of optimized GA-BPNN models.

	Profile Angle (°)	Al Selectivity	DC Bias (V)	Al Etch Rate (Å/min)
Q	0.3	0.5	0.3	0.1
TT	0.0924	0.082	0.1142	0.2668
NHN	9	3	8	7
IWD	2.7511	0.5770	1.4527	1.2300
g_b	1.1943	0.5313	0.5072	0.4200
g_l	0.4460	1.0641	1.1372	1.2800
P-RMSE	1.69	1.70	54	261

As an illustration, the same Al etch rate was optimized. The P-RMSEs and the fitness for the optimized model are shown in Table 7 as a function of Q. As shown in Table 7, the P-RMSE is seen to increase with increasing Q. One optimized model is obtained at 0.1, and the corresponding P-RMSE is 261 Å/min. This shows an improvement of about 39.8% compared to that for the corresponding conventional model. By balancing both errors, the P-RMSE could therefore be considerably improved (reduced).

2. GA-optimized Model

In GA optimization, the experimental ranges for the training factors employed are shown in Table 6. The size of initial population was 100. The crossover and mutation probabilities were 0.9 and 0.1, respectively. The GA evolution was completed at the generation number of 100. The fitness function was defined as

$$\text{Fitness} = \frac{1}{1+\text{PI}} \quad (4)$$

$$\text{PI} = \text{Q} \times \text{T} - \text{RMSE} + (1-\text{Q}) \times \text{P} - \text{RMSE} \quad (5)$$

where T-RMSE represents the model training error. The Q was varied from 0.1 and 0.9 with an increment of 0.2. At each generation, the constructed 200 models were evaluated and only one model with the smallest P-RMSE was selected. This was repeated for 100 generations and one finally optimized model was then determined among 100 selected models.

In the same way, other etch responses were optimized and all results are shown in Table 8. In Table 8, the optimized training factors and P-RMSEs are shown. The percent improvements over the conventional models determined earlier are shown in Fig. 5. As shown in Fig. 5, for the DC bias, no improvement was achieved. However, for the other three etch responses, the improvements are noticeably large. These comparison results strongly support that balancing the training and test errors is effective to improving the GA-BPNN prediction capability.

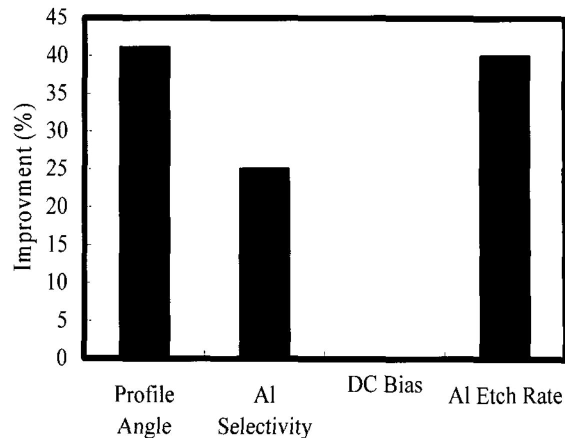


그림 5. 종래의 모델과 GA-BPNN 모델의 예측성능 비교.

Fig. 5. Comparison of prediction performance of conventional and GA-BPNN models.

V. Conclusions

In this study, the effect of training factors on the BPNN prediction performance was optimized by using GA. GA played a role of finding optimized training factor sets as a function of trading factor between the training and test errors. For a systematic modeling, the etch process was characterized by means of a statistical experiment. The presented technique was evaluated with a large number of etch outputs. Also, optimized GA-BPNN models were compared to conventional models. The comparisons revealed that GA-BPNN yielded better predictions in nearly all etch models. This is a strong indicative that controlling the presented technique is effective in improving the BPNN generalization capability. The presented technique is general in that it can be applied to any plasma process data.

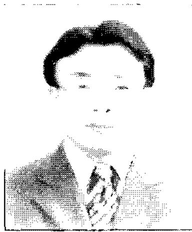
References

[1] C. D. Himmel and G. S. May, "Advantages of plasma etch modeling using neural networks over statistical techniques," *IEEE Trans. Semicond. Manufact.*, vol. 6, no. 2, pp. 103-111, 1993.

[2] S. Bushman, T. F. Edgar, and I. Trachtenberg, "Modeling of plasma etch systems using ordinary least squares, recurrent neural network, and projection to latent structure models," *J. Electrochem. Sco.*, vol. 144, no. 4, pp. 1379-1389, 1997.

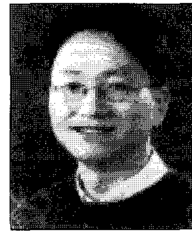
[3] B. Kim, S. Kim, and B. T. Lee, "Modeling SiC surface roughness using neural network and atomic force microscopy,"

J. Vac. Sci. Technol., vol. 22, no. 5, pp. 2467-2472, 2004.
[4] D. E. Rummelhart and J. L. McClelland, *Parallel Distributed Processing*, Cambridge, M.I.T. Press, 1986.
[5] B. Kim and S. Park, "An optimal neural network plasma model: a case study," *Chemometr. Intell. Lab. Syst.*, vol. 56, no. 1, pp. 39-50, 2001.
[6] D. E. Goldberg, *Genetic Algorithms in Search, Optimization & Machine Learning*, Addison Wesley, Reading MA (1989).
[7] B. Kim, S. Y. Lee, and B. T. Lee, "Etching of 4H-SiC in a NF_3/CH_4 inductively coupled plasma," *J. Vac. Sci. Technol. B*, vol. 21, no. 6, pp. 2455-2460, 2003.
[8] D. C. Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons, Singapore, 1991.



김 병 환

1985년 고려대학교 전기공학과(공학사). 1987년 고려대학교 전기공학과(공학석사). 1995년 Georgia Institute of Technology (공학박사). 2001년~세종대학교 전자공학과 부교수 관심분야는 인공지능, 반도체 공학.



우 형 수

1992년 서울대학교 전자공학과(공학사). 1994년 서울대학교 전자공학과(공학석사). 1998년 서울대학교 전자공학과(공학박사). 2001년~세종대학교 전자공학과 부교수 관심분야는 평판디스플레이, 반도체 공학.

곽 관 웅

1993년 고려대학교 기계공학과(공학사). 1995년 고려대학교 기계공학과(공학석사). 2003년 The University of Texas at Austin(공학박사). 2006년~세종대학교 기계공학과 조교수. 관심분야는 비선형 제어, 인공심장 개발, 반도체 열처리 공학.